Evaluating Asymmetric Multicore Systems-on-Chip using Iso-Metrics

Evaluating Asymmetric Multicore Systems-on-Chip using Iso-Metrics

Charalampos Chalios
School of EEECS
Queen's University of Belfast
United Kingdom
cchalios01@qub.ac.uk

Dimitrios S. Nikolopoulos
School of EEECS
Queen's University of Belfast
United Kingdom
d.nikolopoulos@qub.ac.uk

Enrique S. Quintana-Ortí
Depto. Ing. y Ciencia Comp.
Universitat Jaume I, Castellón,
Spain
quintana@uji.es

ABSTRACT

The end of Dennard scaling has pushed power consumption into a first order concern for current systems, on par with performance. As a result, near-threshold voltage computing (NTVC) has been proposed as a potential means to tackle the limited cooling capacity of CMOS technology. Hardware operating in NTV consumes significantly less power, at the cost of lower frequency, and thus reduced performance, as well as increased error rates. In this paper, we investigate if a low-power systems-on-chip, consisting of ARM's asymmetric big.LITTLE technology, can be an alternative to conventional high performance multicore processors in terms of power/energy in an unreliable scenario. For our study, we use the Conjugate Gradient solver, an algorithm representative of the computations performed by a large range of scientific and engineering codes.

Categories and Subject Descriptors

C.1.3 [**Computer Systems Organization**]: Other Architecture Styles—*heterogeneous (hybrid) systems*; G.4 [**Mathematical Software**]: Efficiency

1. INTRODUCTION

The performance of today's computing systems is limited by the end of Dennard scaling [1] and the cooling capacity of CMOS technology [5]. In response, CPU architectures turned towards multicore designs already in the middle of past decade, and power-saving techniques and mechanisms originally conceived for embedded and mobile appliances are being increasingly adopted by desktop and server processors. Near-threshold voltage computing (NTVC) is a promising power-saving technology to tackle the power wall by diminishing voltage (and slightly frequency) of the processor at the cost of reducing hardware reliability [4]. The hope in NTVC is that the (close to) linear drop that is expected in performance from the decay of frequency is compensated by cramming more cores into the same power budget. In addition, the increase in hardware concurrency can be exploited to integrate some sort of algorithmic-based fault tolerance (ABFT) that addresses eventual data corruption caused by operating with unreliable hardware.

In this paper, we investigate the performance, power and energy balance of two representative low power ARM processors of a big.LITTLE system-on-chip (SoC), when applied to a memory-intensive numerical problem. Concretely, our analysis experimentally evaluates the *iso-performance* and *iso-power* of quad-core ARM Cortex-A15 and Cortex-A7 *clusters* against a conventional high performance Intel Xeon E5-2650 CPU, using the Conjugate Gradient (CG) method [6]. This memory-bounded algorithm for the solution of linear systems is particularly interesting as it is representative of the type of operations and performance attained by many other scientific and engineering codes running in high performance computing facilities [2]. As an additional contribution, we shed some light into the energy-saving potential of NTVC under a realistic scenario. For this purpose, we leverage a fault-tolerant variant of CG, enhanced with a self-stabilizing (SS) recovery mechanism [7], to assess the practical energy trade-off between hardware concurrency, CPU frequency, and hardware error rate, using the ARM big.LITTLE architecture as a case study.

As part of related work, iso-energy-efficiency models are built in [8] in order to predict and balance energy and performance in large power-aware clusters, taking into account software characteristics. Compared to this, we focus on the trade-off between performance, power and energy for high-end multicore processors vs low power SoCs, designed mainly for embedded and mobile systems. Our goal is to answer whether it is possible to build systems out of such power-efficient architectures that can match the performance of current throughput-oriented machines. Similarly to us, the authors of [3] study the use of power-efficient architectures in scientific applications. In this line, we take one step further, to make projections about the energy-efficiency of unreliable NTVC platforms and the use of fault tolerance techniques to tackle the unreliability issues.

The rest of the paper is structured as follows. In Section 2 we describe the experimental setup. In Section 3 we compare high performance vs low power architectures using two different iso-metrics, and in Section 4 we determine the effect of unreliable hardware on the CG method. Finally, we close the paper with a few remarks in Section 5.

2. EXPERIMENTAL SETUP

2.1 The CG method

The CG method is a key algorithm for the numerical solution of symmetric positive definite (SPD) sparse and dense linear systems [6] of the form $Ax = b$, where $A \in \mathbb{R}^{n \times n}$ is SPD, $b \in \mathbb{R}^n$ contains the independent terms, and $x^n \in \mathbb{R}^n$ is the solution. The cost of this iterative method is dominated by the matrix-vector multiplication (GEMV) with A that is computed per iteration. For a matrix A with n_z nonzero entries, this operation roughly requires $2n_z$ floating-point arithmetic operations (flops). Additionally, each iteration involves a few vector operations that cost $O(n)$ flops each.

For our evaluation, we employ IEEE 754 real double-precision arithmetic and stop the iteration when the relative residual of the approximated solution is below 1.0e−8. Furthermore, we consider only problems with dense A and, for simplicity, we do not exploit the symmetric structure of the matrix. Under these conditions, we estimate the cost per iteration of CG to be $2n^2$ flops (i.e., we neglect the lower cost of the vector operations). Moreover, for efficiency, we leverage multi-threaded implementations of the GEMV kernel in Intel MKL (version 11) for the Intel-based CPU, and ATLAS (version 3.8.4) for the ARM-based cores.

2.2 Target architectures and scenarios

The experiments in this paper were performed using three different CPUs. The first one, hereafter XEON, is a high-performance but power-hungry Intel Xeon E5-2650 socket with 16 GBytes of DDR3-1333 MHz RAM. The alternative low-power architectures, A15 and A7, are two ARM quad-core clusters embedded into an Exynos5 system-on-chip (SoC) of an ODROID-XU board, sharing 2 Gbytes of DDR3-800 MHz RAM. Table 1 offers the most important features of these CPU architectures. There, the "Stream bandwidth" column reports the memory bandwidth measured using the `triad` test of the `stream` benchmark[1] on the highest number of cores available in the sockets. The "Roofline GFLOPS" column corresponds to the theoretical upper bound on the computational performance (in terms of GFLOPS, or billions of flops per second) dictated by the *roofline model*.

For the evaluation, we investigate different scenarios that vary in the number of cores (from 1 up to the maximum), the CPU frequency, and the problem size. For simplicity, we only consider two CPU frequencies (lowest and highest, in particular discarding Intel's turbo-mode) for each architecture; and two problem dimensions: an "on-chip" case that occupies much of the last level of cache (LLC), n=1,024 on XEON, n=512 on A15 and n=256 on A7; and an "off-chip" counterpart that clearly exceeds the capacity of the LLC, with n=4,096 on XEON, n=1,024 on A15 and n=512 on A7.

3. HIGH PERFORMANCE VS LOW POWER

In this section, we perform an experimental evaluation of the target CPU architectures, using the CG method (implemented on top of optimized multi-threaded versions of MKL and ATLAS), from the points of view of performance, power dissipation, and energy consumption. The purpose of this analysis is to expose the trade-offs between these three metrics, for a memory-bound method such as CG, on these particular architectures, with the ultimate goal of answering two key questions:

- Q1 (*Iso-performance*): Can we attain the performance of the Intel Xeon CPU with the low power ARM clusters while yielding a more power-efficient solution?
- Q2 (*Iso-power*): What is the performance that can be attained using the low power ARM clusters within the power budget dictated by the Intel Xeon socket?

3.1 Trade-offs

Figure 1 reports the results from the evaluation of the multi-threaded CG implementations, from the points of view of performance (in GFLOPS), power dissipation (W) and energy efficiency (GFLOPS/W), using both on-chip and off-chip problems. We note that an evaluation in terms of GFLOPS and GFLOPS/W allows a comparison of these metrics for problems of varying size, which require a different number of flops.

We start by distinguishing between the two scenarios corresponding to on-chip and off-chip problems. For brevity, we will focus hereafter in the former case, noting that, in the latter, the performance on XEON and A15 is clearly limited by the memory bandwidth, offering considerably lower figures on all three metrics. The same memory bottleneck is not visible for A7 though, likely because the multi-threaded implementation of the matrix-vector multiplication in ATLAS does not extract all the performance of this architecture.

Table 2 offers numerical results for the on-chip problems. Our comments to these results are organized in three axis: #cores, frequency and architecture (configuration parameters) as well as three perspectives (metrics). Let us commence by putting the light on the #cores. From the point of view of concurrency, increasing #cores produces fair speed-ups, which interestingly are quite close for all three architectures independently of their frequency; e.g., the use of 4 cores on XEON, A15 and A7 produces speed-ups between 2.8 and 3.4 for any of the two frequencies. From the perspective of power, a linear regression fit to the data shows a high value of the y-intercept for XEON, which basically corresponds to static power, and can be explained by its large LLC, the complex pipeline, the large area dedicated to branch prediction, etc. Compared with this, A15 and A7 exhibit much lower static power, reflecting the simpler design of this CPU clusters. This difference between the Intel- and ARM-based architectures has a major impact on the energy where, e.g., increasing the #cores on XEON results in shorter execution time and, due to the large static power, a visible positive effect on energy efficiency (GFLOPS/W). This is a clear indicator of the potential benefits of a "race-to-idle" policy applied to this architecture. The effect of increasing #cores on A15 and A7 is more imprecise, due to the low fraction that the static power represents.

We continue next with the analysis of frequency. Independently of the number of cores, the effect of this parameter on performance is perfectly linear for XEON but sublinear for A15, where doubling the frequency only improves performance by a factor of about 1.7×; and slightly higher for

[1] `http://www.cs.virginia.edu/stream`

Acron.	CPU socket/cluster	#Cores	Frequency range (GHz)	LLC: level, type, size (Mbytes)	TDP (W)	Peak mem. bandwidth (GBytes/s)	Stream mem. bandwidth (Gbytes/s)	Roofline GFLOPS
XEON	Intel Xeon E5-2650	8	1.2–2.0	L3, shared, 20	95	51.2	44	11
A15	ARM Cortex-A15	4	0.8–1.6	L2, shared, 2	N/A	N/A	5.4	1.35
A7	ARM Cortex-A7	4	0.5–1.2	L2, shared, 0.5	N/A	N/A	2.07	0.51

Table 1: Hardware specifications of the target architectures.

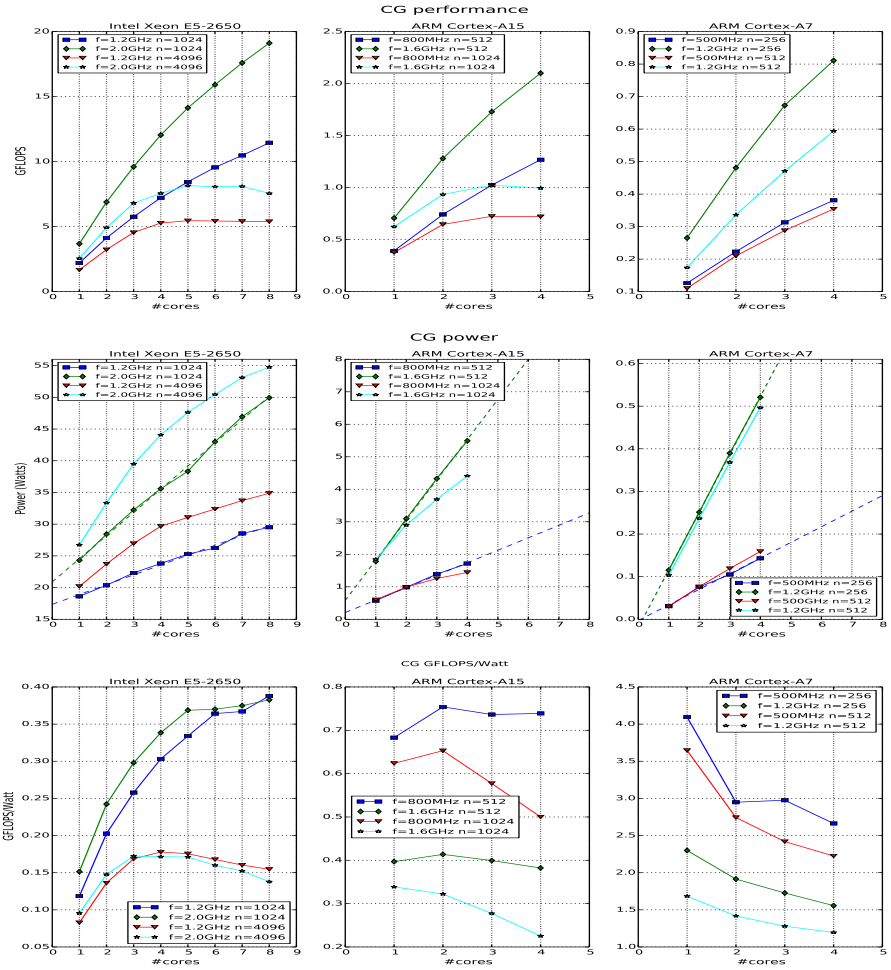

Figure 1: Evaluation of performance, power and energy on the target architectures using multi-threaded implementations of the CG method on both the on-chip and off-chip problems.

A7, where raising the frequency from 0.5 to 1.2 GHz (a factor of 2.4×) results in an increase of performance 2.1×. The effect of frequency on power is sublinear for XEON (a factor between 1.30–1.69×, depending on the number of cores) and superlinear for both A15 (3.12–3.20×) and A7 (3.66–3.71×). The net effect of the variations of time and power with the frequency is that, on XEON, increasing the frequency slightly improves energy efficiency (race-to-idle) while on the ARM-based clusters it reduces it by a factor close to 50% for A15 and 64% for A7.

Finally, we observe some general differences between the CPU architectures: the power hungry 8-core Intel CPU produces significantly higher performance rates (and, therefore, shorter execution times) than the ARM clusters, at the expense of a much higher dissipation rate and lower energy efficiency. The differences between A15 and A7 follow a similar pattern, with higher performance in the former in exchange for higher power draft/lower energy efficiency.

3.2 Analysis of iso-metrics

We open the following study by noting that the questions Q1 (iso-performance) and Q2 (iso-power) formulated at the beginning of this section can be analyzed in a different number of configurations/scenarios. Here we select one that we find specially appealing. Concretely, for Q1 we consider the

CPU	Freq. (GHz)	#cores	Time per iter. (ms)	Performance (GFLOPS)	Speed-up	Power (W)	Energy (GFLOPS/W)
Xeon	1.2	1	0.89	2.21	1.0	18.6	0.12
		2	0.47	4.12	1.9	20.4	0.20
		4	0.27	7.21	3.3	23.8	0.30
		6	0.21	9.55	4.3	26.2	0.36
		8	0.17	11.44	5.2	29.5	0.39
	2.0	1	0.53	3.67	1.0	24.3	0.15
		2	0.28	6.88	1.9	28.4	0.24
		4	0.16	12.04	3.3	35.5	0.34
		6	0.12	15.91	4.3	42.9	0.37
		8	0.10	19.11	5.2	49.9	0.38
A15	0.8	1	1.26	0.39	1.0	0.57	0.68
		2	0.66	0.74	1.9	0.98	0.75
		4	0.39	1.26	3.2	1.71	0.74
	1.6	1	0.70	0.70	1.0	1.78	0.40
		2	0.40	1.28	1.8	3.09	0.41
		4	0.25	2.10	2.8	5.49	0.38
A7	0.5	1	0.98	0.12	1.0	0.03	4.09
		2	0.56	0.22	1.8	0.07	2.95
		4	0.32	0.38	3.0	0.14	2.66
	1.2	1	0.48	0.26	1.0	0.11	2.30
		2	0.26	0.48	1.2	0.25	1.91
		4	0.16	0.81	2.9	0.52	1.56

Table 2: Evaluation of performance, power and energy on the target architectures using multi-threaded implementations of the CG method on the on-chip problems.

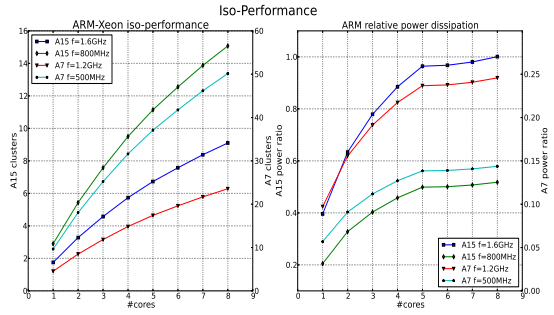

Figure 2: Evaluation of iso-performance. Left: Number of A15 or A7 clusters o match the performance of a given number of Xeon cores at 2.0 GHz. Right: Comparison of power rates dissipated for configurations delivering the same performance.

performance of 1–8 cores from Xeon, at 2.0 GHz, as the objective, and then we evaluate how many clusters (consisting of A15 or A7 and operating at either the lowest or the highest frequencies) are necessary to match the reference performance. Question Q2 is the iso-power counterpart of the Q1, with the power budget reference fixed by the power rate of 1–8 cores from Xeon, at 2.0 GHz. Because of the scalability issue, in all cases we employ the performance and power rates observed when operating with on-chip problems.

The left-hand side plot in Figure 2 reports the results from the iso-performance study, exposing that, in order to attain the performance of 8 cores from Xeon (2.0 GHz), it is necessary to use about 9.1 A15 clusters (i.e., quad-cores) at 1.6 GHz or more than 50.2 A7 clusters at 0.5 GHz! (Note the different scales of the y-axis depending on the type of cluster). Now, we recognize that in such comparison we implicitly introduce a simplifying assumption in favour of the ARM CPUs. In particular, for the on-chip problem on Xeon, the dimension n=1,024. Now, in order to solve the same problem on a multi-socket ARM platform, data and operations have to be partitioned among and mapped to the clusters, incurring into overhead due to communication. For the CG method, we can expect that this additional cost comes mostly from the reduction vector operations (analogous to a synchronization). Also, there is a certain overhead due to operating with a smaller problem size per core.

The right-hand side plot in Figure 2 illustrates the ratio between the power rates dissipated by four configuration "pairs" that attain the same performance, with one of the components of these pairs being Xeon and the other A15 or A7, at either the lowest or the highest frequency. Following with the previous examples, 8 cores from Xeon (at 2.0 GHz) deliver the same performance as 9.1 clusters from A15 at 1.6 GHz, and they draw basically the same power rate (a ratio of 1.001 between the two). On the other hand, using 50.2 clusters of A7 at 0.5 GHz only requires a fraction of the power rate dissipated by Xeon, concretely 14%.

Figure 3 displays the results from the complementary study on iso-power. The plot in the right illustrates that with the power budget of 1–8 Xeon cores, it is possible to accommodate a moderate number of A15 clusters or a very large volume of A7 ones. The performance ratio between these ARM-based clusters with respect to the Xeon, in the left plot, reveals decreasing gains with the number of A15 clusters and a performance tie with respect to 4 or more Xeon cores. The ratio also decays for the A7 clusters, but in this case it is stabilized around a factor of 7.

Note that not all ARM-based configurations considered in the iso-performance and iso-power study have the same on-chip memory capacity (iso-capacity) as Xeon. In particular, given that the LLC for the latter is 20 MBytes, one need at least 10 A15 clusters and 40 A7 clusters to be in an iso-

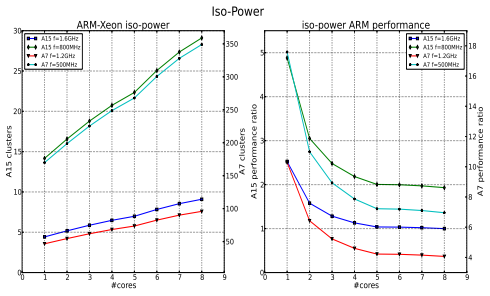

Figure 3: Evaluation of iso-power. Left: Number of A15 or A7 clusters that match the power dissipated by a given number of XEON cores at 2.0 GHz. Right: Comparison of performance rates attained for configurations dissipating the same power rate.

capacity scenario from the on-chip memory point of view.

We conclude this section by noting that a study of the energy efficiency ratio under the conditions imposed by Q1 or Q2 does not contribute new information. For example, given that Q1 basically relates the GFLOPS/W of two architectures with equal GFLOPS rates, an evaluation of energy efficiency boils down to the analysis of the power ratio.

4. ENERGY COST OF RELIABILITY

The experiments and analysis in this section aim to expose the potential impact on energy exerted by a technique that, like NTVC, trades off lower CPU (voltage and) frequency and, therefore, more reduced power consumption, for increased hardware concurrency and failure rate. In order to perform this study in a realistic scenario, we raise the following considerations:

- We employ a tuned variant of our multi-threaded implementations of the CG method, equipped with a SS recovery mechanism [7] to cope with silent data corruption introduced by unreliable hardware. Following the experiments in [7], the SS part is activated every 10 iterations of the CG method, and must be performed in reliable mode. From the computational point of view, the major difference between an SS iteration and a "normal" CG one is that the former performs a total of two GEMV instead of only one. However, these two GEMV can be performed simultaneously, as they both involve A. Therefore, for a memory-bound operation like GEMV, we can consider that in practice, the two types of iterations share the same computational cost.
- To accommodate a reliable+unreliable execution, we consider an "ideal" multi-socket big.LITTLE SoC consisting of a single quad-core A15 cluster plus several A7 clusters. Here, A15 operates at the highest frequency, is considered to be reliable, and applies the SS mechanism. On the other hand, the A7 clusters operate at the lowest frequency, represent the unreliable hardware, and are used to compute the normal CG iterations. We will refer to this SoC as A15+NA7, and we will use data corresponding to on-chip problems for all the experimentation.
- The convergence rate of the CG iteration depends on the condition number of matrix A [6]. Under certain conditions, the convergence of the SS variant degrades logarithmically with the error rate [7]. Silent data corruption is assumed to occur during GEMV, producing one or more bit flips into any of its results, and propagates from there to the rest of the computations. The convergence rate of the SS variant also depends mildly on whether the bit flips are bounded to the sign/mantissa or can affect also the exponent.

Case study	A15+NA7 #A7 clusters	A15+NA7 GFLOPS	A15+NA7 Power
iso-performance	5.51	2.09	1.24
iso-power	38.85	13.49	5.44
iso-capacity	4	1.57	1.05

Table 3: Comparison of A15+NA7 to A15 under iso-performance, iso-power and iso-capacity conditions.

Under these conditions, we next perform an experimental analysis of the energy gains that such a reliable.unreliable big.LITTLE SoC features, comparing it with a reliable single quad-core A15 cluster operating at the highest frequency under iso-performance and iso-power conditions.

We commence with the iso-performance study. The first goal is to find how many A7 clusters must be involved during the execution of the CG iterations so that, when combined to build A15+NA7 with a single A15 cluster for the execution of SS iterations (10% of the total), the performance that is obtained matches that of a single A15 cluster operating at the highest frequency (i.e., 2.1 GFLOPS; see Table 2). A little arithmetic gives an answer of 5.51 A7 clusters, which we will round to 6 A7 clusters, at the price of attaining a performance slightly above the reference objective (concretely, 2.28 GFLOPS). We can next compare the power dissipation rate of the two cases: 5.49 W for A15 and 1.31 W for A15+NA7. Next, the GFLOPS rates for each two configurations, combined with the cost per iteration ($2n^2$) and the number of iterations required for convergence in the n=512 case, offers the execution times (slightly smaller for A15+NA7, because of the rounding). A combination of time with the previous power rates thus offers the *energy-to-solution* (ETS), i.e., how much energy (in Joules) is required to solve the same problem, on each architecture, in absence of errors (though A15+NA7 applies the SS mechanism nonetheless). Finally, in Figure 4, we compare the ETS attained by original CG method, executed in a reliable environment, against that of the SS variant, under unreliable conditions, as the convergence degrades a certain percentage of iterations due to errors. These results explicitly expose the energy gains that can be expected from operating with simpler low power cores, at low frequencies, for this particular application, with A15+NA7 outperforming A15 in terms of ETS when the degradation incurs in up to 340% more iterations.

We also perform an analogous study from the point of view of iso-power; that is, we set the power dissipated by the A15 cluster, at the highest frequency, as the reference (5.49 W; see Table 2), and then we derive how many A7 clusters can be embedded into A15+NA7 within the same power bud-

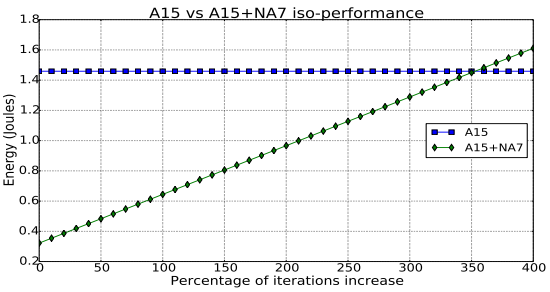

Figure 4: Iso-performance ETS for the original CG method executed by A15 at the highest frequency (reliable mode) and the SS variant of CG executed by A15+nA7 under unreliable conditions which degrade convergence.

get, with the answer being 38.85. This exercise will, eventually, produce the same ETS as the iso-performance analysis. This is to be expected, since any increase of #A7 clusters in A15+nA7 yields an proportional increase of its GFLOPS rate, or equivalently an inversely proportional decrease in execution time. Simultaneously, the power dissipation will be increased in the same proportion, yielding the same ETS.

To conclude this section, we focus on the iso-capacity problem. For this case-study, we require the aggregated LLC of the A7 clusters in A15+nA7 to be equal that of A15. Figure 1 shows that it is important that the data involved in the computation fit in the LLC so that the performance will scale with #cores. Now, A15 includes a 2MB LLC cache, which can hold a problem size of n=512 for CG. Therefore, four A7 clusters match the LLC capacity of a single A15 cluster (see Table 1). In conclusion, we can build an A15+nA7 system which can solve the same problem size as A15, with a throughput of 1.57 GFLOPS, i.e. 1.33× slower than A15, but dissipates 5.22× less power. The iso-performance, iso-power and iso-capacity results are summarized in Table 3.

5. CONCLUSIONS AND FUTURE WORK

The requirement for energy-efficient systems on the road towards Exascale systems asks for more power-efficient hardware designs. In this paper, we turn to the embedded and mobile world and investigate whether platforms from that domain can be used to build systems for HPC applications with better energy-to-performance ratios. Concretely, we show that, in principle, it is possible to use power-efficient ARM clusters in order to match the performance of a high-end Intel Xeon processor while operating, in a worst-case scenario, at the same power budget. Conversely, it is also possible to use a rather large number of ARM clusters, fit into the power budget of one Intel Xeon processor, and attain higher performance.

As a second contribution, we experiment with a reliable CG execution in an A15 cluster versus an execution of a self-stabilizing variant of this method using a hybrid configuration of A15 +A7 to emulate an unreliable processor that operates close to NTV. From this study, we found that one can improve ETS even when the errors slow down the convergence of CG up to 340%.

As cornerstone of CG method is the matrix-vector product, we believe that the significance of this study carries over to many other numerical methods for scientific and engineering applications. On the other hand, the study has certain limitations. For example, we did not consider factors such as the cache hierarchy, interconnection networks, memory buses and bandwidth, which can be significant in large-scale designs and affect both performance and power consumption. We made this choice in order to be able to extract some first-order conclusions about the potential of employing NTVC, and we intend to investigate those matters in more depth in the future.

Acknowledgments

E.S. Quintana-Ortí was supported by project TIN2011-23283 of the MINECO and FEDER, and the EU project FP7 318793 "EXA2GREEN". This work was partially done while this author was visiting Queen's University of Belfast. We thank F.D. Igual, from *Universidad Complutense de Madrid*, for his help with the Odroid board.

This research has been supported in part by the European Commission under grant agreements FP7-323872 (SCoRPiO), FP6-610509 (NanoStreams) and by the UK Engineering and Physical Sciences Research Council under grant agreements EP/L000055/1 (ALEA), EP/L004232/1 (ENPOWER) and EP/K017594/1 (GEMSCLAIM)

6. REFERENCES

[1] R.H. Dennard, F.H. Gaensslen, V.L. Rideout, E. Bassous, and A.R. LeBlanc. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE J. Solid-State Circuits*, 9(5):256–268, 1974.

[2] K. Asanovic et al. The landscape of parallel computing research: A view from Berkeley. Technical Report UCB/EECS-2006-183, University of California at Berkeley, EECS, 2006.

[3] D. Göddeke, D. Komatitsch, M. Geveler, D. Ribbrock, N. Rajovic, N. Puzovic, and A. Ramirez. Energy efficiency vs. performance of the numerical solution of PDEs: An application study on a low-power ARM-based cluster. *J. Computational Physics*, 237(0):132–150, 2013.

[4] U.R. Karpuzcu, Nam Sung Kim, and J. Torrellas. Coping with parametric variation at near-threshold voltages. *Micro, IEEE*, 33(4):6–14, 2013.

[5] R. Lucas *et al.* Top ten Exascale research challenges, 2014. `http://science.energy.gov/~/media/ascr/ascac/pdf/meetings/20140210/Top10reportFEB14.pdf`.

[6] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[7] P. Sao and R. Vuduc. Self-stabilizing iterative solvers. In *Workshop Latest Advances in Scalable Algorithms for Large-Scale Systems*, pages 4:1–4:8, 2013.

[8] S. Song, C. Su, R. Ge, A. Vishnu, and K. Cameron. Iso-energy-efficiency: An approach to power-constrai ned parallel computation. In *IEEE Int. Parallel Distr. Proc. Symp. (IPDPS)*, pages 128–139, 2011.